

Analysis of Errors in Machine Translation from Roger T. Bell's Translation Process Model

Jianbin Zhu¹, Min Zhang^{1, a}

¹College of Foreign Studies, Guilin University of Technology, Guangxi 541004, China

^a2242455535@qq.com

Abstract

There are enough literature reviews about machine translation, but the numbers of texts studied are not large enough, and there are very limited varieties of machine translation tools they have used. Taking the Chinese Government Work Report in 2020 as an example, this paper compares Google, Youdao, Baidu, Sogou and official translations from the perspective of syntax and pragmatic structure by using Roger T. Bell's translation process model. It is found that new words are not introduced into the Frequent Lexis Store in time, and the organization of sentence structure is not appropriate, and pragmatic awareness is weak through the process of machine translation. In the same text, Baidu translation has the most words, and Sogou translation has fewer grammatical problems, and Youdao translation has better pragmatic performance, and Google translation has no outstanding highlights.

Keywords

Machine translation; Translation process, syntax; Pragmatics; Translation of external publicity.

1. Introduction

Scholars have made endless researches on errors of machine translation, but it is not systematic for researches of errors in machine translation from the perspective of the translation process. Roger T. Bell put forward a complete model of translation process from semantic representation in his book *Translation and Translation Process: Theory and Practice*. Based on this theory, this paper will discuss how machine translation makes mistakes through the process of translation, with a view to help translators make better revisions after translation, at the same time this paper would also give translators guidelines to use suitable tools of translation when they are burdened with too many tasks and various orders of translation .

2. Literature Review

Using computer technology to analyze, process and study human natural language began in 1950s (Zhiwei Feng, 2019). Early machine translation is Phrase-Based Machine Translation (PBMT). In 1990s, China launched Statistical Machine Translation (SMT) of large-scale real corpora. At the beginning of the 21st century, scientific research institutions and enterprises represented by Google have successively set up development teams of Statistical Machine Translation. A few years later, Baidu and other companies in China have successively released Internet machine translation systems that can support dozens of common languages in the world. Around 2014, with the development of computer deep learning technology, Neural Machine Translation (NMT) appeared, which significantly improved the quality of translation. At present, Chinese companies such as Baidu, Sogou, Youdao have adopted the method of Neural Machine Translation (Mu Li, 2019). The translation works produced by the same method are also various in their expressions. Although it's hard to say that there is a standard

theory about the quality of translated works, there is better one among a lot of translated works. Rejwanul Haque, Mohammed Hasanuzzaman and Andy Way compared PBMT and NMT in terms translation between English and Hindi. In their research, they used 2000 sentence pairs in judicial fields to form test sets, and annotated the test sets by marking terms in the source language and target language of the test sets, so as to establish a golden standard for measuring terminology translation errors. Therefore, there are rules to measure the quality of translated works. Errors of machine translation can be roughly classified as: missing translation, wrong translation, incorrect form, inconsistent style, improper word order and punctuation errors (Iacer Calixto&Qun Liu,2019). Scholars have focused on translation results before, but in recent years, more and more scholars have paid attention to translation process. According to CNKI (a database collection system with many academic journals, papers and so on), in the visual analysis of document retrieval with the theme of “translation process”, the number of such papers reached 2,361 in 2016, and even if it has declined in recent years, it is expected that there will be 1,781 related papers in 2020. In addition, the methods of studying the translation process are constantly being updated. For example, the Think Aloud Method, Keystroke Method, Eye Tracking Research, Neuroscience Method (Fan Yang, Defeng Li, 2018), Questionnaire Method (Wenzhao He, Defeng Li, 2018). These studies are more inclined to empirical study of translation process, while studies of machine translation are more and more inclined to quantitative study. However, the number of texts and tools of translation studied are too limited. Therefore, this paper combines quantitative research with qualitative research, uses Bell’s translation process model theory, compares and analyzes four versions of machine translation, and looks back at the machine translation process from the translation results.

3. Roger T. Bell’s Translation Process Model

Bell used language, psychology and information theory to describe the translation process, thus revealing the translation knowledge and translation methods that translators should master (Ping Gou, Hao Wang, 2013). Bell’s book *Translation and Translation: Theory and Practice* can be divided into three parts, which are model, meaning and memory respectively. This article mainly talks about the first part: mode. Bell regards the translation process as a top-down decoding process and a bottom-up encoding process. There are three stages involved in decoding. First: syntactic stage. The translator needs to start from the source text, takes clauses as the unit, makes the Visual Word Recognition System work, and searches for frequent structures and words that may exist in the Frequent Structure Store and Frequent Lexis Store in the source text. If the search fails, at this time, it’s a necessity to ask help from Parser and Lexical Search Mechanism, which will start to search for related structures and words in the long-term memory and short-term memory of human brain. Second: semantic stage. In the semantic stage, the translator mainly deals with transitivity, which refers to the material process, mental process, verbal process, behavioral process, existential process and relational process mentioned in Systemic Functional Grammar. Third: pragmatic stage. At this stage, the translator mainly deals with the thematic structure and register of clauses, which includes three parameters: tenor, mode and field. After three stages of the analysis process, the translator’s cognitive process will reach the Idea Organizer and Planner. If the Planner decides to carry out the translation task at this time, the translation process will enter the encoding process, and the order of which is just opposite to that of decoding process, including pragmatic stage, semantic stage and syntactic stage as its order. The translation process won’t end itself until the target language is formed. But as Bell said, this process is not a linear process, but a coordinated integration process (Roger T. Bell, 1991), see Fig. 1.

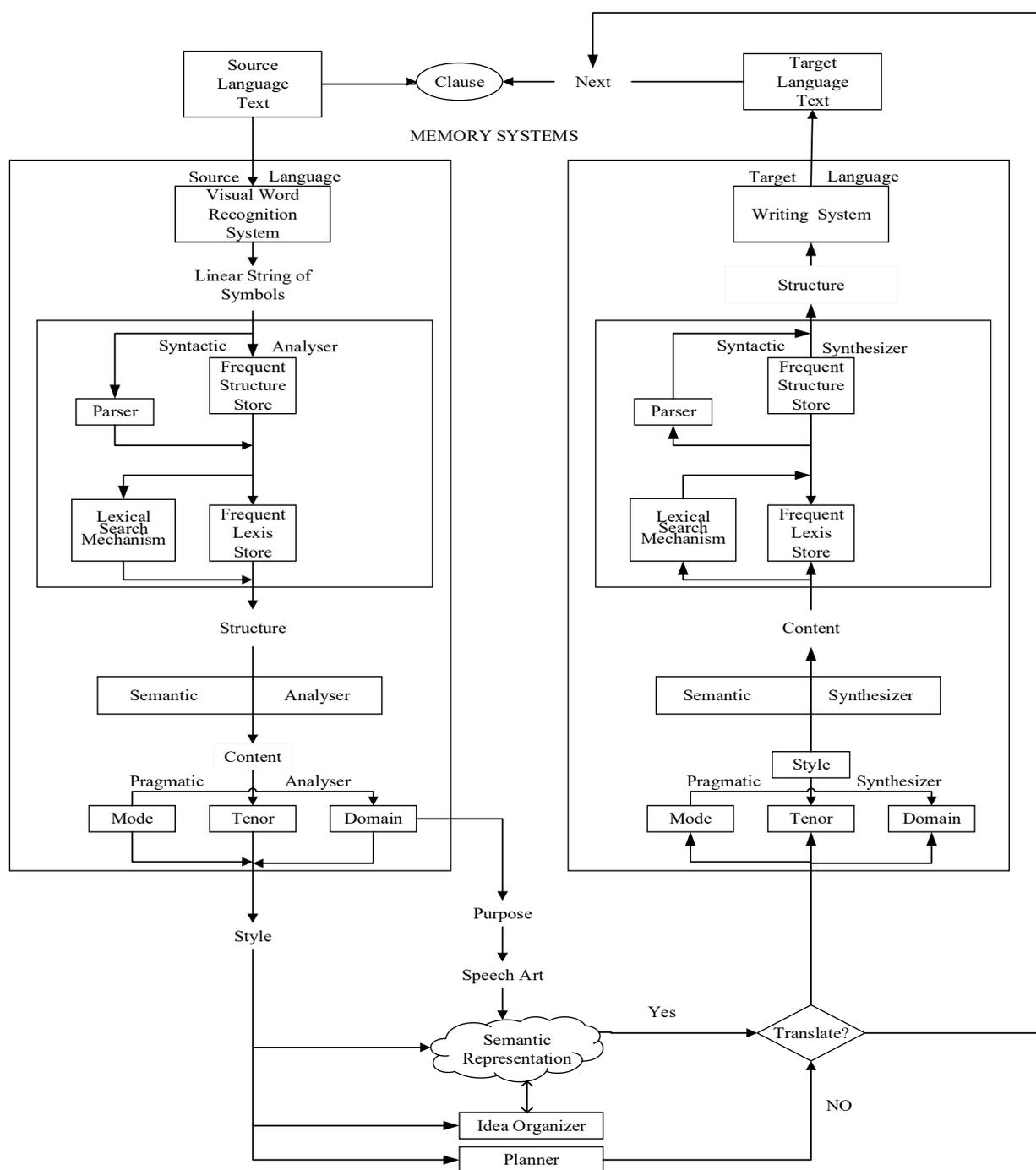


Fig 1. Roger T. Bell's model of translation process

4. The Application of Roger T. Bell's Translation Process Model in the Process of Machine Translation

4.1. Syntactic Stage

4.1.1. Lexical Level

On the syntactic level, Bell mainly analyzes the frequent lexis and structures, which are the priorities should be considered when we translate source language. Judging from the number of words in the translation, there are 11,344 words in the the Report in Chinese and 9,611 words in the official translation. There are both omissions in the four kinds of machine translations, but the degree is different. The specific data see Table 1.

Table 1 Comparison of the number of words between different versions of translation

Version	Chinese	Official	Baidu	Youdao	Google	Sogou
Number of Words	11344	9611	8663	8493	8279	8025
Number of Words Less Than Official One	/	/	948	1118	1332	1586

By comparing five different versions, this paper finds that one of reasons for omission lie in the literal and inaccurate translation of proprietary words. For example, when translating “Adhere to Four Concentrations”, four versions are as follows:

Official text: concentrated patients, experts and resources in designated facilities to provide treatment

Youdao: adhere to the "four"

Google/Baidu:adhere to the "four concentration"

Sogou:adhere to the "four concentrations"

The “Four Concentrations” mentioned above mainly refer to the concentration of “patients, experts, resources and facilities” to defeat covid-19, but machine translation has not given the actual connotation. In the above versions, Google and Baidu have a few grammatical errors, but Youdao has more serious grammatical error. Thus, Sogou did better than others.

In addition to phrase translation, we can also make further analysis through sentences as follows:

Official text: “... see that every cent is used where it is needed most and where market entities and the people will feel the greatest benefit from it.”

Youdao:We must make sure that every sum of money is spent where it matters most. We must make sure that market entities and the people have the real feeling.

Google:...every money must be spent on the cutting edge and important points, and the market entities and the people must have a real feeling.

Sogou:...every sum of money must be used at the cutting edge and where it is urgent, so that market players and the people must have real feelings.

Baidu:...spend every sum of money on the edge of the knife and at the most important places, and let the main body of the market and the people have a real feeling.

The first one is the official translation, but when four kinds of machine translation software translate “where it is needed most”, only Youdao adopts omission, and other three kinds of machine translation software are literal translations, which the translation of Baidu even is “on the edge of the knife”, which is illogical. As for the sentence “the people will feel the greatest benefit from it”, none of the four machine translation software can translate the real meaning. In addition, the word “so that” appeared in the translated version of Sogou, which regards the first sentence as a causality of the second sentence, and does not translate them into compound sentence like that of Google and Baidu. This shows that Sogou has considered the relationship between sentences in the translation process. As a result, Sogou has done a better job in this respect, followed by Youdao.

From the perspective of Bell’s translation process, part of the reason of such omission caused by machine is that these special words or latest words are not stored in the Frequent Lexis Store by machines, which makes them unable to successfully extract related words in the translation process, and still stays at the primary stage of just putting words together, so they fail to translate the actual meaning of the source language, which leads to translation failure. This shows that the technology of machine translation has not been improved to keep pace with the times.

In addition to the problem of omission, the four kinds of machine translation also have different degrees of wrong translation. In 2019, due to raged covid-19, the word “covid-19” frequently showed up in the Report with a total number of 41 cases. The most suitable word to refer to this kind of epidemic happened across the globe in 2019 is “covid-19” (Jiangxia Zeng, 2020). From the lexical level, word frequency and collocation are important indicators to measure the quality of machine translation (Qing Wang, Xiao Ma, 2020). By comparison, five different versions translates the word “covid-19” in different ways, see Table 2.

Table 2. Words related to “covid-19” appeared in five versions

Version	Official	Baidu	Youdao	Google	Sogou
Word					
“covid-19”	25 times	0 time	7 times	0 time	3 times
“epidemic”	6 times	31 times	24 times	28 times	33 times
“pandemic”	2 times	0 time	0 time	0 time	0 time
Grand Total	27 times	31 time	31 times	28 times	36 times

According to the data, the word “epidemic” is mainly selected by the four tools when translating word “covid-19”. In Baidu and Google, we can not see the word “covid-19”, while Youdao and Sogou own it in 7 times and 3 times respectively. On Wikipedia, the word “epidemic” refers to “an epidemic is the quick spread of a kind of disease to a lot of people in a given population within a short time” and the explanation of “covid-19” is “an infectious disease closely related to the SARS virus. It is mainly spread between people by small droplets from infected persons when they breathe or cough.” According to their different definitions, “covid-19” is not equivalent to “epidemic”, but a kind of “epidemic”. Although Baidu and Google have records of translating “covid-19” (Jiangxia Zeng, 2020), they did not accurately translate the word “covid-19” in the Report. Therefore, Baidu and Google did not successfully extract the frequent lexis when translating, which made it difficult to meet the current needs. From this point of view, Youdao and Sogou are better at machine translation than Baidu and Google.

4.1.2. Structural Level

Table 3. Main verbs in the sentences without subject in *the Report*

The type of verbs	Total times of occurrence
“promote”	25 times
“strengthen”	21 times
“support”	12 times
“deepen/optimize”	12 times
“improve”	11 times
“implement”	10 times
“expand”	9 times
“ensure”	9 times
“develop”	8 times
“accelerate”	6 times
“increase”	5 times
“advance”	5 times
“adhere to”	5 times

English belongs to Indo-European language family, which pays more attention on the organized form and structure with obvious subject-predicate-object structure and more inanimate

sentences, while Chinese belongs to Sino-Tibetan language family, which emphasizes harmonization in meaning and has more sentences without subject. To some extent, this also reflects the differences of thinking and speaking habits between Chinese and western people. The differences between English and Chinese also cause trouble for machine through the process of translation. As a kind of publicity document, the Report contains a large number of statements without subject. According to statistics, the most common sentences without subject in the Report begin with verbs, and some of the verbs that appear the most are shown in the Table 3.

According to the principle, when translating Chinese sentences without subject, translators can usually adopt methods of adding subjects or translating them into passive sentence or translating them into imperative sentences or translating them into the structure of “it+ formal subject” or translating them as the structure of “there be...” and other methods (Man Wang, 2000).

According to statistics, the five different versions have different methods in dealing with sentences without subject. The official translation adopts the method of adding subject “we” accompanied with a few passive sentences. In machine translation, Youdao mainly adopts these two methods, but supplemented by a few imperative sentences beginning with verbs. Baidu takes the main method of adding subject, and there are obviously more imperative sentences and passive sentences. Google and Sogou translate most sentences without subject into imperative sentences that begin with verb. The data of how four tools of machine translation translate the phrase “we will” that is the most frequent in the official version when translating sentences without subject sees in Table 4.

Table 4. Numbers of the phrases “we will” occurred in different versions

Version	Official	Youdao	Baidu	Google	Sogou
Number of the phrase “we will” occurred	211 times	199 times	107 times	12 times	6 times

In order to further understand the translation of sentence without subject, this paper selects an example from the Report, and the results of translation are as follows:

Official text: “We have kept up with the spring farming schedule, and continued the critical battle against poverty.”

Youdao: Do not miss the spring ploughing. We will work tirelessly to alleviate poverty.

Google: Do not miss the farming season and catch spring ploughing. Unremittingly promote poverty alleviation.

Sogou: Don’t miss the farming season and catch the spring ploughing. Make unremitting efforts to get rid of poverty.

Baidu: Do not miss the farming season, pay attention to spring ploughing. We will make unremitting efforts to overcome poverty.

There are many short sentences in Chinese, so when translating this kind of sentences, we need to connect them together according to their internal meaning. But only the official version adopts the method of combining sentence, and four tools of machine translation divide sentences just according to full stops. Therefore, the translator can use the method of subdividing sentences to make the machines do their job better (Qing Wang, Xiao Ma, 2020). Judging from the translation of sentences without subject, Youdao and Baidu are better than others.

4.2. Pragmatic Stage

The pragmatic level mentioned by Bell includes three parameters: field, tenor and mode. Field refers to what actually happens, including the environment in which language occurs and the topic being discussed (Xiaotang Cheng, 2002). Field of the text studied in this paper is the Report Li Keqiang delivered at the Chinese National People's Congress. The content is simply summarized as "looking back on the past and looking forward to the future", which points out the direction of future work.

Tenor refers to the relationship between participants, including their social status and role relationship. In the conference, Li Keqiang, as the premier of China, is also a "public servant of the people", and his report was shown to the deputies of the National People's Congress and even the entire Chinese people. When the translator translates the Report into English, the audience will turn to foreigners who understand English.

Mode is the medium or channel of language communication. The language of the Report takes into account both rigorous and formal features of written language and smooth and concise features of spoken language. There are many declarative sentences and imperative sentences in the Report, while declarative sentences are used to narrate the truth and imperative sentences are used to suggest things (Lou Zhixin, 2005).

There are 13 structures of "Yao+verb" (means we need to...) in the Report. Comparing the five versions of translation, there are a lot of imperative sentences in Sogou and Google's translations. Official, Youdao and Baidu both translate imperative sentences into sentences with the subject "we". On the surface, Sogou and Google's translations are closer and superior to the original style. However, from a deeper or pragmatic perspective, the result is not the case. Imperative sentences express command, prohibition, suggestion, dissuasion, request and begging for exemption, and their tone is rather blunt. There are many imperative sentences in the Report. The reasons lie in the facts that firstly, there are many sentences without subjects in Chinese, and there will be no ambiguity without stating the subject; Secondly, because of covid-19, the previous Report with more than 20,000 words only has more than 10,000 words this year, and the time of report has been reduced to less than one hour (Yingjie Liu, 2020), so the language must be more concise. However, as an external publicity document, the Report aims to convey China's policies and thoughts and establish China's image. Attention should be paid to the differences of languages and cultures between Chinese and English. If translated into imperative sentences, it ignores the fact that China is a socialist country, and the Communist Party of China aims to serve the people wholeheartedly. Imperative sentences can not highlight the collectivist thought—"we". Imperative sentences add the atmosphere that the Chinese Communist Party gives its orders, while sentences with the structure of "we+verbs" reflects the pragmatic effect that the implementation of future policies will be worked together by the whole Chinese people under the leadership of the Communist Party of China, making the people feel more cordial. From this point of view, the latter is superior than the former. Therefore, from the perspective of structure and pragmatics, Youdao and Baidu's translations are more in line with the needs.

5. Conclusion

By analyzing the syntactic and pragmatic aspects of four versions translated by machines, we find that even if Neural Translation System is adopted, there are still problems in machine translation, such as omission, wrong translation and format mistake. If Bell's theory is used to explain these problems, the main reason is that some new words are not introduced into the frequent lexis store of machine in time, and the sentence structure is not perfect, which leads to the failure of the extraction through the process of translation and the weak pragmatic awareness of machine translation. As far as machine translation itself is concerned, in the same

text translated, Baidu owns more words, Sogou has fewer grammatical problems, and Youdao is better in pragmatic level, but more serious errors of format, and Google has no outstanding highlights. After being aware of their peculiar characteristics of such tools, translators can turn to different tools of translation according to their different needs. Compared with previous studies, this paper has made some progress in the number of texts, but the number of texts still needs to be increased. In addition, machine translation itself has some errors in the internal operation of the system. For example, even if the same source language and the same tool of machine translation is used, there will be subtle differences in the target language translated twice. Therefore, machine translation still needs further research.

Notes:

The original text of “The Chinese Government Work Report” is selected from the Central People’s Government Network of the People’s Republic of China. The website can be found in <http://www.gov.cn/guowuyuan/zfgzbg.htm>

The official translation is selected from the official English website of the Central People’s Government of the People’s Republic of China and People’s Daily Online.

http://english.www.gov.cn/premier/news/202005/30/content_WS5ed197f3c6d0b3f0e94990da.html

Acknowledgements

This work is supported by the 2019 Innovation Project of Guangxi Graduate Education (Grant No. JGY2019099)

References

- [1] Roger T. Bell: Translation and Translating: Theory and Practice (Longman Singapore Publishers, British, 1991).
- [2] Rejwanul Haque, Mohammed Hasanuzzaman, Andy Way: Analysing Terminology Translation Errors in Statistical and Neural Machine Translation, Machine Translation, Vol.67 (2020) No.34, p. 2-3.
- [3] Iacer Calixto, Qun Liu: An Error Analysis for Image-based Multi-modal Neural Machine Translation, Machine Translation, Vol.67 (2019) No.33, p. 1-2.
- [4] Xiaotang Cheng: Register Theory and Semantic and Pragmatic Analysis of Poetry, Foreign Language and Foreign Language Teaching, Vol.42 (2002) No.4, p. 11-14.
- [5] Xinjie Cai, Bing Wen: Statistical Analysis of Error Types in Chinese-English Machine Translation—Taking Chinese-English Translation of Foreign Propaganda Texts as an Example, Journal of Zhejiang Sci-Tech University (Social Science Edition), Vol.42 (2020) No.4, p. 1-8.
- [6] Yingying Deng, Bin Wang: A Review of Translation Process Research from Cognitive Perspective, English Square, Vol.21 (2019) No.105, p. 22-23.
- [7] Ping Gou, Hao Wang: A Review of Translation Process Research Based on Cognitive Psychology, Journal of Sichuan University of Technology (Social Science Edition), Vol.35 (2013) No.3, p. 50-54.
- [8] Wenzhao He, Li Defeng: Evaluation of Chinese Translation of English Relative Clauses taking Google Machine Translation as an Example, China Science and Technology Translation, Vol.33 (2019) No.32, p. 30-34.
- [9] Zhixin Lou: Interpreting the Linguistic Features of Premier Wen Jiabao’s Government Work Report, Journal of Zhejiang Ocean University (Humanities Edition), Vol.39 (2005) No.3, p. 15-19.

- [10] Yingjie Liu: Features, Emphases and Highlights of the Government Work Report, China Development Watch, Vol.16 (2020) No.11, p. 5-7.
- [11] Danling Li: Cognitive Process of Irony Translation the Conception and Practice of Combining Bell's Translation Process Model with Schema Theory, Journal of Guangdong Polytechnic Normal University (social sciences), Vol.41 (2014) No.6, p. 106-112.
- [12] Zhiwei Feng: 70 Years of Computational Linguistics Research in China, Language Education, Vol.34 (2019) No.4, p.19-29.
- [13] Ran Tao: Literature Review of Translation Process Research, Journal of Foreign Languages College of Jinling Institute of Technology (Social Science Edition), Vol.36 (2012) No.4, p. 84-89.
- [14] Yifang Wang: Eye Tracking Data Collection, Presentation and Analysis in Translation Process Research, Foreign Language Research, Vol.37 (2017) No.34, p. 76-81.
- [15] Qing Wang, Xiao Ma: Research on the Quality Evaluation Method of Machine Translation from the Perspective of Problem Consciousness, Hunan Social Sciences, Vol.33 (2020) No.6, p. 134-141.
- [16] Manliang Wang. English Translation Principles of Chinese Sentences without Subjects, Foreign Language Teaching, Vol.64 (2000) No.2, p. 66-69.
- [17] Fan Yang, Defeng Li: Research Methods of Translation Process: Problems and Prospects, Chinese Science and Technology Translation, Vol.33 (2018) No.31, p. 1-5.
- [18] Jiangxia Zeng: Analysis of Translation Criteria of Scientific and Technological Terms in the Context of Multimodal Big Data—Taking COVID-19 Pneumonia and SARS-COV-2 Terminology Translation as Examples, Chinese Science and Technology Terminology, Vol.36 (2020) No.22, p. 33-41+71.
- [19] Meifang Zhang, Fei Ke: Roger T. Bell's Perspective of Linguistic Translation Studies, Foreign Languages and Translation, Vol.21 (2001) No.3, p. 11-16.
- [20] <http://www.gov.cn/guowuyuan/zfgzbg.htm>
- [21] http://english.www.gov.cn/premier/news/202005/30/content_WS5ed197f3c6d0b3f0e94990da.html